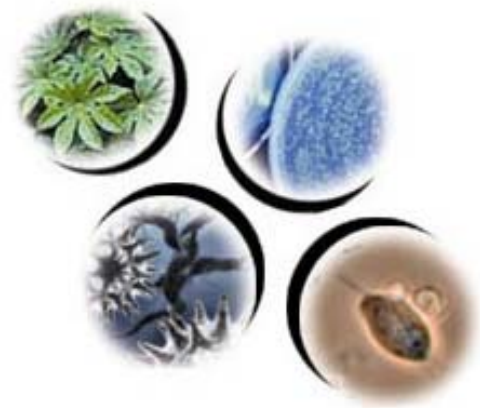
BioHealthBase

BioDefense Public Health Database



BioHealthBase Development Approach

NIH Contract Number: HHSNS66200400041C ADB Contract Number: N01-AI-40041

May 17, 2005

Presenter: Carey Gire

UT SOUTHWESTERN MEDICAL CENTER

NATIONAL INSTITUTES OF HEALTH

NIAID

NORTHROP GRUMMAN Information Technology

Partnership

BioHealthBase
BioDefense Public Health Database

Development Goals and Challenges

- Goals
  - Be flexible as new science requirements arise
    - Provide rapid turnaround on new features
  - Build a system that supports the science that researchers need
    - Collaborate with researchers during analysis and development
  - Integrate existing science software & schemas wherever possible
- Challenges
  - Scientific priorities change frequently
  - Each scientist has a different perspective on science needs
  - Science software & schemas are diverse and independent

BioHealthBase
BioDefense Public Health Database

Agile Development Approach

- Project Management with Scrum
  - Iterative, incremental, evolutionary development
  - Short development windows ("Sprints")
    - 2 – 6 weeks (nominally, 30 calendar days)
  - Maintain a "Project Backlog" of candidate tasks
  - Plan & prioritize features at the beginning of each Sprint
  - Focus on prioritized features during the Sprint
  - Daily Scrum status meeting – report status, identify issues for off-line resolution
  - Always have a working system
- Test-driven Development/Continuous Integration
  - Develop test software prior to functional software
  - Daily software builds,"unit" tests, and integration tests

BioHealthBase
BioDefense Public Health Database

BioHealthBase Analysis Approach

- Use Case Development
  - Identified standard use cases from bioinformatics data centers
  - Refined and augmented with UTSW and other science feedback
- Researcher Survey
  - PubMed search for researchers affiliated with our organisms
  - Broadcast survey; received 250 responses
  - Results on the BHB website:
    - http://www.biohealthbase.org/whats_new.htm
- Integrate Science Community Feedback
  - Began meeting with organism teams and plan for more
  - Intend to involve interested science stakeholders in prototype evaluations, system testing, data analysis, etc.

BioHealthBase
BioDefense Public Health Database

Scientific Questions Summary

- Host-pathogen interactions (bacterial proteins – host proteins/pathways)
  - Proteins that contribute directly to disease pathogenesis – virulence factors
    - Basic mechanisms, therapeutic targets (symptoms), drug resistance, diagnostic targets
  - Proteins that support the co-existence of host and pathogen – survival factors
    - Basic mechanisms, therapeutic targets (eradication), drug resistance
  - Proteins that serve as host immune system targets
    - Basic mechanisms of immune response to infection, active vaccine development, passive vaccine development
  - Proteins that influence host range/transmission
- Population dynamics and evolution
- Regulation of gene expression
- Other Items of Interest
  - Training, tutorials, FAQs
  - Epidemiology/PublicHealth issues

BioHealthBase
BioDefense Public Health Database

BioHealthBase Development Approach

- Results of Requirements Analysis
  - Documented requirements and issues
  - Identified requirements "gaps" → potential prototypes
- Rapid Prototyping – This Summer
  - Started with GUS system; also looking at GMOD/Chado
    - Plan to load Francisella, influenza, M. tb, and supporting data
    - Integrate BioCyc metabolism pathways and signaling pathways
    - Begin exploration of epidemiological data sources
    - Begin evidence collection to support ongoing annotation
  - Plan to deploy a variety of tools for prototype evaluation and feedback
    - Genomic browsers, alignment, annotation, queries, and others
  - Begin building some infrastructure frameworks (APIs, etc.)

BioHealthBase
BioDefense Public Health Database

What Have We Learned?

- GUS
  - Broad schema, good content, reasonable set of load software
  - GUS wdk is complex at times – would like to simplify the question plug-in model and database access layer
- GMOD
  - Excellent tools for many specific functions
  - Would like to see better encapsulation of database independence in bioperl layer
- What Would We Like to See?
  - Standard database encapsulation layer – GUS wdk DTO? EJB? DAO?
  - Ability to pipe results of a tool into another tool
  - Ability to easily integrate plug-ins or other add-ons into existing tools
  - Standard process to extend schemas – epidemiology, pathways, etc.
  - Build frameworks that:
    - Assist BRC tool & data sharing
    - But also facilitate community participation in tool development and extended data models

BioHealthBase
BioDefense Public Health Database

BioHealthBase Plan of Attack

- Get Feedback on Approaches of Other Development Teams
- Plan to Collaborate with Existing Development Teams
  - GUS team – schema and wdk enhancements, more tool integrations, wiki documentation
  - GMOD team – tool enhancements, new tools
  - Other BRC teams – tool and data sharing
- Design BioHealthBase to Be “Community-Built”
  - We want to create frameworks that facilitate collaboration
    - E.g., identify useful APIs that users can build on top of – and encourage them to build their own tools
  - The IOWG effort is key

BioHealthBase
BioDefense Public Health Database

Goals and Challenges - Revisited

- Encourage “Good” System Development
  - Build the “flexibility to change” into our development approach
  - Find interested researchers and put ‘em to work
  - Integrate creatively; develop frugally
  - Collaborate with existing development teams
  - Build an infrastructure framework that allows collaboration on the development of the system